JOURNAL OF VIROLOGY, Oct. 1985, p. 7–11
0022-538X/85/100007-05$02.00/0

Vol. 56, No. 1

Nucleotide Sequence of a Human Cytomegalovirus DNA Fragment Encoding a 67-Kilodalton Phosphorylated Viral Protein

MICHELLE GALL DAVIS[1]* AND ENG-SHANG HUANG[1,2]

Lineberger Cancer Research Center[1] and Department of Medicine and Department of Microbiology and Immunology,[2] University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514

Received 11 February 1985/Accepted 10 June 1985

Previously, we demonstrated that a DNA fragment from the L-unique region at 0.37 to 0.39 map units of human cytomegalovirus DNA codes for a 67-kilodalton phosphorylated tegument protein. (Davis et al., J. Virol. 52:129–135, 1984). We have determined the nucleotide sequence of this DNA, transcription initiation and termination, and putative promoter region. We also compared the properties of the protein product predicted from the DNA sequence with the properties of the protein isolated by immunoabsorbent column chromatography. This viral structural protein may have important functions, such as protein kinase activity, DNA binding, and possible transcriptional activation of immediate-early genes.

Human cytomegalovirus (HCMV) is a herpesvirus with cascading transcription, as described by several investigators (3, 11). One of the strongest regions of transcription late in infection is located in the center of the L-unique region (2). A DNA fragment which includes this transcription region was transferred to simian virus 40-transformed monkey kidney cells (COS-1) and a 67-kilodalton (67K) cytomegalovirus protein was detected in extracts of these cells by reaction of blotted proteins with virus-specific monoclonal antibody (2). The 67-K protein is a phosphorylated protein found in isolated virions. It probably corresponds to the 66K phosphorylated tegument protein described by Gibson (4) in the simian cytomegalovirus system.

The 67-K protein has been isolated by immunoabsorbent column chromatography (E.-C. Mar and E.-S. Huang, manuscript in preparation). It binds to DNA and phosphorylates proteins, an activity that was previously discovered in isolated virions (7) and in infected cells (9). In this communication, we present the nucleotide sequence of the coding region and mapping of the RNA transcript.

MATERIALS AND METHODS

Recombinant plasmids. Plasmid pHD7 (pBR322-HCMV *Eco*RI-G) and recombinant M13-713 were previously described (2). The 2.9-kilobase (kb) *Bgl*II fragment from plasmid pHD7 was inserted at the *Bam*HI site of the replicative-form DNA of M13mp7 to generate recombinant M13-713. The *Bgl*II-digested DNA of plasmid pHD7 and *Eco*RI-digested, replicative-form DNA of M13-713 were used for hybridizations and sequence analysis.

Sequence analysis. DNA was labeled for nucleotide sequence analysis by incorporation of α-^{32}P-labeled deoxynucleotide monophosphate at the 3′ termini with large-fragment polymerase (Bethesda Research Laboratories) or by incorporation of [γ-^{32}P]ATP at the 5′ termini with polynucleotide kinase. The chemical degradation method of Maxam and Gilbert (8) was used. A summary of labeled fragments is illustrated in Fig. 1. Enzymes were obtained from Bethesda Research Laboratories and digestions were carried out as recommended by the supplier.

RNA mapping. End-labeled DNA fragments prepared for sequencing were also used as primers for cDNA synthesis. DNA primers were hybridized to RNA isolated from WI-38 cells infected with the Towne strain of HCMV and extended by reverse transcriptase to determine the location of the 5′ end of the RNA transcript. RNA was isolated from infected cells by lysis with 7 M guanidine hydrochloride and Sarkosyl followed by centrifugation through 6 M CsCl at 34,000 rpm for 12 h in an SW41 rotor. RNA pellets were extracted with phenol and chloroform. Approximately 100 μg of total RNA was mixed with 3′ labeled DNA fragments. The mixture (25 μl) was heated at 95°C for 5 min and slowly cooled to 40°C. In a final volume of 100 μl, the RNA-DNA hybrids were incubated with 16 mM KCl–50 mM Tris hydrochloride (pH 8.3)–10 mM $MgCl_2$–0.1 mM four deoxynucleoside triphosphates–10 mM β-mercaptoethanol–100 μg of bovine serum albumin per ml–3 μl of reverse transcriptase (Life Sciences, Inc.). After incubation at 37°C for 3 h, samples were extracted with phenol. RNA was degraded by addition of 0.1 M NaOH and incubation for 1 h at room temperature, followed by addition of 0.15 M NaH_2PO_4. Nucleic acids were repeatedly ethanol precipitated to remove salts, suspended in 80% formamide, denatured by boiling, and loaded onto 4% acrylamide gels containing 7 M urea. cDNA was detected by autoradiography.

Total cellular RNA was also isolated from infected cells labeled with $^{32}P_i$ as described previously (2).

Nucleic acid hybridization. Nick translations and hybridizations were performed as described previously (2).

RESULTS

Nucleotide sequence of the 67K gene region. Enzymes *Hae*III, *Hin*fI, *Dde*I, *Alu*I, and *Taq*I were used to fragment the isolated 2.9-kb *Eco*RI fragment of M13-713 replicative-form DNA. These fragments were end labeled and chemically degraded. The sequence listed in Fig. 2 begins at the first determined nucleotide, 165 base pairs (bp) upstream of the leftward *Bgl*II site of Fig. 1, at 0.37 map unit. This sequence and the nucleotides 106 bp downstream of the rightward *Bgl*II site at 0.39 map unit were determined by labeling *Bgl*II fragments of plasmid pHD7 DNA. The orientation of these *Bgl*II fragments on the HCMV map was determined by hybridization to pHD7 DNA immobilized on nitrocellulose filters. The entire sequence, 3,072 bp, is listed in Fig. 2.

RNA mapping. A 239-bp fragment from bases 717 to 956

* Corresponding author.

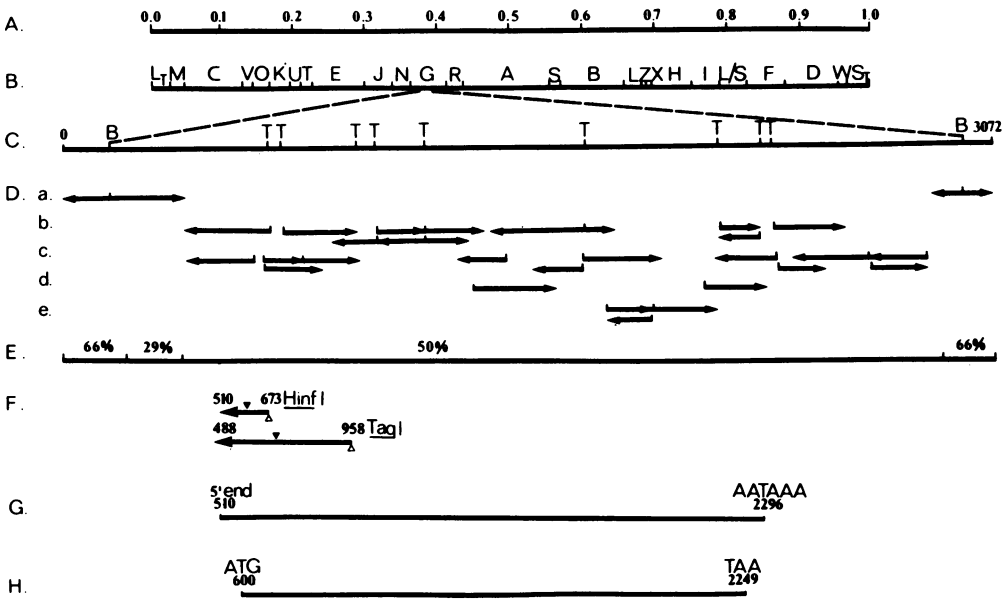

FIG. 1. Physical map of the HCMV DNA molecule. (A) Fractional length of the HCMV DNA molecule. (B) Recognition sites for *Eco*RI endonuclease on DNA isolated from HCMV Towne strain, passage 36, previously published (1). (C) Expanded region of fragment *Eco*RI-G, indicating the extent of the nucleotide sequence. 0, First nucleotide determined; 3072, last nucleotide determined. T, Recognition sites for *Taq*I endonuclease. B, Recognition sites for *Bgl*II endonuclease. (D) Labeled fragments (dots) and nucleotide sequence determinations (arrows) used to compile the sequence. Fragments labeled at: (a) *Bgl*II sites; (b) *Taq*I sites; (c) *Hin*fI sites; (d) *Dde*I sites; (e) *Hae*III sites. (E) G+C content of the DNA. (F) Fragments labeled at the 3′ end (▼) and extended by reverse transcriptase (arrows): △, nucleotide corresponding to the 5′ end of the priming fragment; ◀, nucleotide corresponding to the 3′ end of the cDNA product. (G) Boundaries of the RNA transcript. (H) Boundaries of the protein translation region.

generated by *Taq*I digestion of M13-713 DNA was labeled by incorporation of [α-^{32}P]dCMP at the 3′ ends. After extension of this primer with reverse transcriptase, a cDNA product 470 bases long was detected (Fig. 3). This observation was used to estimate the location of the RNA initiation at base 488 (5′ *Taq*I site at 958-cDNA product 470). A 62-bp fragment from bases 608 to 670 generated by *Hin*fI digestion of M13-713 DNA was labeled by incorporation by [α-^{32}P]dAMP. After extension of this primer with reverse transcriptase, a cDNA product 164 bases long was detected. This product was compared with *Hin*fI-digested pBR322 and a partially degraded DNA fragment 230 bases long. The 3′ end of this cDNA is at base 510, 164 bases from the 5′ *Hin*fI site at base 673, corresponding to the 5′ end of the RNA. See Fig. 1 for a schematic illustration of this experiment.

The 3′ end of the RNA is probably near base 2295, at sequence AATAAA. This is indicated by the hybridization of RNA to the fragmented DNA of plasmid pHD7 (Fig. 4). Digestion of DNA of plasmid pHD7 with enzyme *Taq*I generates an 1,150-bp fragment which includes the sequence between bases 0 and 668 (the 5′ *Taq*I site is not included on the sequence), a doublet at 510 bp which includes sequences between bases 1025 and 2087, and a 760-bp fragment which includes sequences between bases 2304 and 3072 (the 3′ *Taq*I site is not indicated on the sequence, although it must be close to the last determined nucleotide). The 1,150-bp fragment and the doublet at 510 bp hybridize with RNA isolated from infected cells, but the 760-bp fragment does not (Fig. 4). Thus, the RNA termination is 5′ to base 2304, probably at base 2296, sequence AATAAA.

Protein coding. The initiation codon AUG occurs at base 548, again in phase at base 566, and the terminator TAG at base 605. We have no information about possible in vivo synthesis of this 19-amino acid oligopeptide.

The codon AUG also occurs at base 600 and the coding region continues for 549 amino acids to base 2246. The composition of the protein encoded by this sequence is very basic: 20% lysine, arginine, and histidine, 5.5% aspartic and glutamic acids, 36% hydrophobic residues, and 38% other residues. The 549-amino acid protein would be 66 kilodaltons, close to the size predicted by gel electrophoresis of the protein immunoprecipitated by monoclonal antibodies specific for this gene product. The basic composition of the protein is consistent with DNA binding activity.

DISCUSSION

HCMV DNA was mapped by electron microscopy of partially denatured DNA molecules (5), illustrating that some low guanine-plus-cytosine (G+C) stretches are easily found and other regions of high G+C composition are also present. The first 250 bp and the final 250 bp of this 3,072-bp stretch listed in Fig. 2 contain 66% G+C bases, but the sequence 250 to 500 bp, just upstream from the 5′ end of the RNA, contains 29% G+C bases, and the sequence between 500 and 2840 bp, including the entire coding region, contains 50% G+C bases (Fig. 1E).

The initiation of RNA transcription at base 510 in the sequence listed in Fig. 2 was determined by primer extension of DNA fragments which end near the 5′ end of the RNA. Total RNA was used as a template; this RNA contains a single 1.9-kb species which hybridizes with the primer fragments (2). Since the only radioactive label added was the priming fragment, the only labeled product was the cDNA resulting from extension of the primer. Indeed, the major product in two separate experiments was a single band with few intermediate or obviously degraded products (Fig. 3).

The sequences often observed 5′ to the initiation of RNA transcription in eucaryotes are observed at typical locations

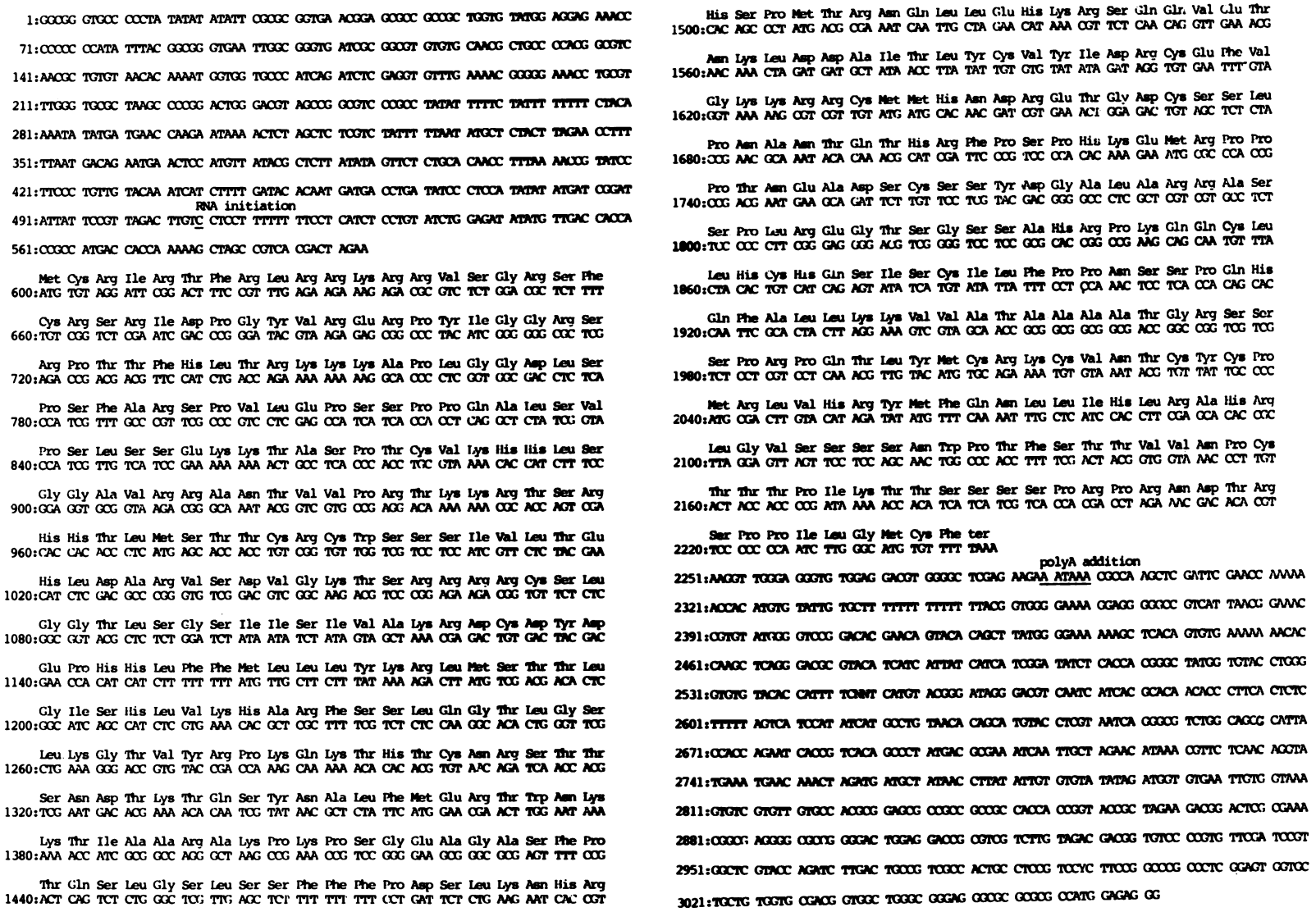

FIG. 2. Nucleotide sequence of HCMV DNA, 0.37 to 0.39 map units. The sequence of the anticoding strand is written 5′ to 3′. The interpreted amino acid sequence is indicated in the open reading frame from base 600 to 2249. Underlined bases are the 5′ end of the RNA at base 510 and the polyadenylation site at base 2295.

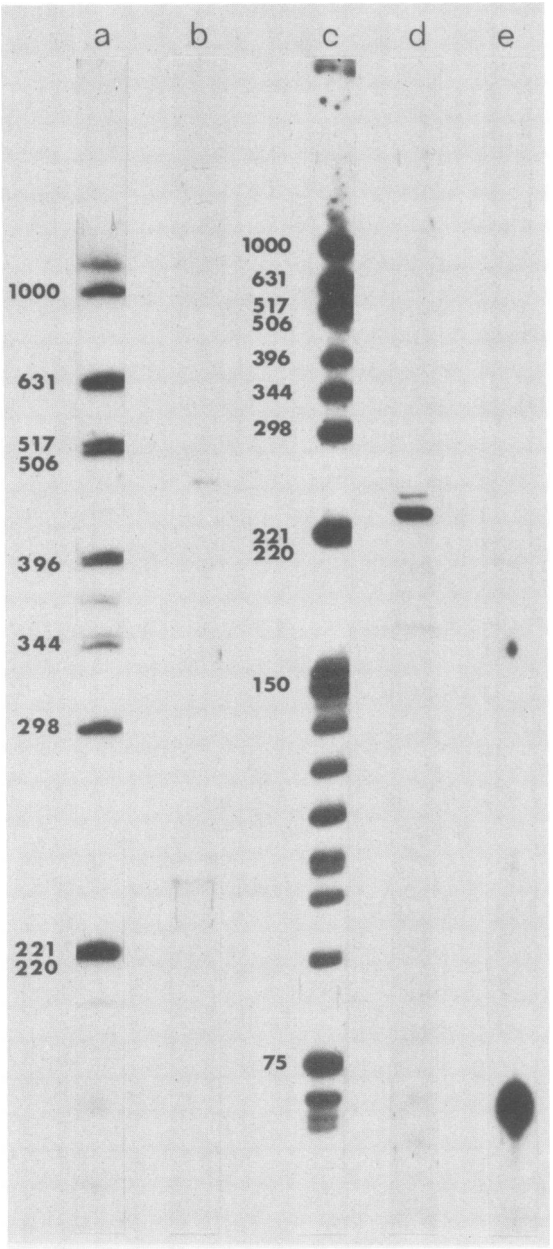

FIG. 3. Primer extension of 3′ end-labeled DNA fragments. X-ray exposures of 4% acrylamide gels containing fragments of pBR322 plasmid DNA digested with *Eco*RI and *Hin*fI, labeled with [α-^{32}P]dAMP and large-fragment polymerase (lanes a and c), cDNA primed with a 239-bp fragment (lane b), a partially degraded DNA fragment (lane d), and cDNA primed with a 70-bp fragment (lane e). Additional fragments in lane c correspond to M13mp7 DNA. The fragment length in bases for lane a is indicated to the left, and the fragment length in bases for lane c is indicated to the right.

5′ to the initiation of transcription at base 510 of this sequence. The sequence TATATATATA occurs at base 24 (486 bases upstream), ATATATG occurs at base 283 (228 bases upstream), TATATAG occurs at base 385 (126 bases upstream), CAAT occurs at base 452 (58 bases upstream), and ATATATATG occurs at base 475 (45 bases upstream). Some alignment of related sequences can be drawn which includes pyrimidine tracts and alternating AT sequences; examples are illustrated in Fig. 5. Many hairpin loop structures can be drawn which involve these sequences; the significance of such structures is unknown. The direct repeats reported by Stinski et al. (10) were not found, and no long direct repeats were detected by computer analysis. This DNA region is only transcribed in infected cells after the initiation of viral DNA synthesis, and WI-38 cells treated with recombinant plasmids including this coding region do not produce the 67K protein. We presume that an α or β viral gene product is required to stimulate expression of this viral gene.

The termination of the RNA transcript near base 2295 is indicated by the sequence AATAAA. The RNA isolated from infected cells does not hybridize with the 760-bp DNA fragment that includes the sequence from base 2304 to base 3072, generated by *Taq*I digestion of pHD7 DNA. Thus, the

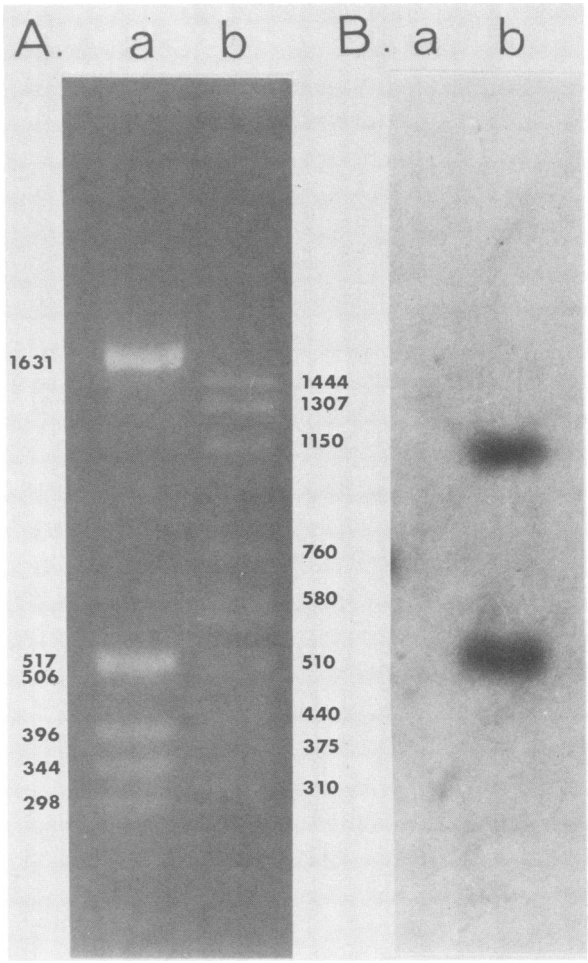

FIG. 4. Hybridization of HCMV RNA to DNA of plasmid pHD7. A 2% agarose gel (A) containing DNA fragments of plasmid pBR322 digested with *Hin*fI (lane a) and plasmid pHD7 digested with *Taq*I (lane b) was blotted onto nitrocellulose paper which was hybridized with ^{32}P-labeled RNA isolated from WI-38 cells infected with HCMV. Fragment lengths in lane a are 1,631, 517, 506 (doublet), 396, 344, and 298 bp. Smaller fragments are smeared on this gel. Fragments corresponding to pBR322 sequences in lane b are 1,444, 1,307, 475, 368, and 312 bp. Many of the stained bands smaller than 1 kb contain several comigrating species. (B) An X-ray exposure of the nitrocellulose paper after hybridization. Hybridizing fragments correspond to a 1,150-bp fragment and a 510-bp doublet in lane b.

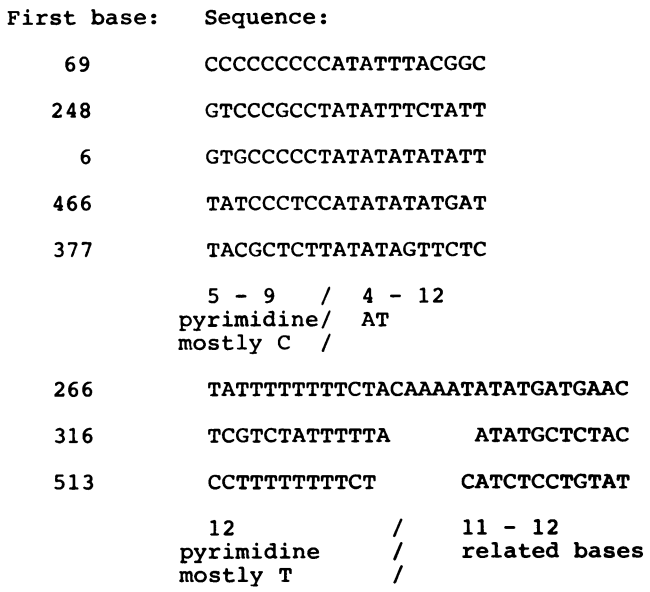

FIG. 5. Alignment of repeated sequences occurring 5′ to the initiation of transcription.

RNA must terminate 5′ to base 2304. There are no apparent splices in the sequence, and our RNA-DNA heteroduplexes observed by electron microscopy do not contain any "R loops," indicating no large spliced regions. The mRNA from infected cells which hybridizes with this DNA is 1.9 kb (2), which would include 1,785 bases from bases 510 to 2295, plus 100 A bases at the 3′ end.

The protein-coding region indicated by the sequence is sufficient for the protein that we have determined is coded by this DNA fragment. We have no amino acid sequence data on this protein product. The monoclonal antibody which identifies the 67K protein in extracts of monkey kidney cells carrying the DNA was used to isolate the 67K protein by immunoabsorbent column chromatography. The protein kinase activity eluted with the 67K protein (Mar and Huang, in preparation). Michelson et al. (9) also reported that the protein kinase isolated from HCMV-infected cells was 68 kilodaltons.

We do not detect transcription of this coding region in HCMV-infected cells incubated in cycloheximide, although the 67K protein is observed in infected cells after uptake of the virus. We believe the protein is taken into the cell along with the viral DNA. It is a basic protein with protein kinase activity, it binds to DNA, and it may be the structural component which functions as an initiator or key enzyme to phosphorylate an initiator of immediate-early gene transcription (10). A similar function has been suggested for the major tegument protein of herpes simplex virus (1).

Protein kinase activity has been measured in isolated HCMV virions (7) and herpes simplex virus virions (6). In HCMV, the protein kinase is apparently a tegument protein (2). The ratio of virus particles/PFU is estimated to be more than 500:1, and those defective virus particles are packed with tegument proteins: thus, thousands of molecules of protein kinase are transferred to each infected cell.

HCMV-infected cells carry on a significant level of host cell protein synthesis, while herpes simplex virus-infected cells do not. Perhaps an enzyme function, such as the viral protein kinase, carried into the infected cells will alter the activity of the host cell machinery. A comparison of the protein kinase functions coded by these viruses may provide some reason for this difference.

ACKNOWLEDGMENTS

This study was supported by Public Health Service grants CA21773 and CA19014 from the National Cancer Institute and grant AI15036 from the National Institute of Allergy and Infectious Diseases.

LITERATURE CITED

1. **Campbell, M. E. M., J. W. Palfreyman, and C. M. Preston.** 1984. Identification of herpes simplex virus DNA sequences which encode a *trans*-acting polypeptide responsible for stimulation of immediate-early transcription. J. Mol. Biol. **180:**1–19.
2. **Davis, M. G., E.-C. Mar, Y.-M. Wu, and E.-S. Huang.** 1984. Mappng and expression of a human cytomegalovirus major viral protein. J. Virol. **52:**129–135.
3. **Demarchi, J. M.** 1981. Human cytomegalovirus DNA: restriction enzyme cleavage maps and map locations for immediate-early, early, and late RNAs. Virology **114:**23–38.
4. **Gibson, W.** 1981. Structural and nonstructural proteins of strain Colburn cytomeglovirus. Virology **111:**516–537.
5. **Kilpatrick, B. A., and E.-S. Huang.** Human cytomegalovirus genome: partial denaturation map and organization of genome sequences. J. Virol. **24:**261–276.
6. **Lemaster, S., and B. Roizman.** 1980. Herpes simplex virus phosphoproteins. II. Characterization of the virion protein kinase and of the polypeptides phosphorylated in the virion. J. Virol. **35:**798–811.
7. **Mar, E.-C., P. C. Patel, and E.-S. Huang.** 1981. Human cytomegalovirus-associated DNA polymerase and protein kinase activities. J. Gen. Virol. **57:**149–156.
8. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**49–560.
9. **Michelson, S., M. Tardy-Panit, and O. Barzu.** 1984. Properties of a human cytomegalovirus-induced protein kinase. Virology **134:**259–268.
10. **Stinski, M. F., R. M. Stenberg, and W. F. Goins** 1984. Structure and function of the human cytomegalovirus immediate early genes, p. 399–421. *In* F. Rapp (ed.), Herpesvirus. Alan R. Liss, Inc., New York.
11. **Wathen, M. W., and M. F. Stinski.** 1982. Temporal patterns of human cytomegalovirus transcription: mapping the viral RNAs synthesized at immediate early, early, and late times after infection. J. Virol. **41:**462–477.